

Lawrence Berkeley National Laboratory

Recent Work

Title

Linear-scaling algorithm for rapid computation of inelastic transitions in the presence of multiple electron scattering

Permalink

<https://escholarship.org/uc/item/18w8s03d>

Journal

Physical Review Research, 1(3)

ISSN

2643-1564

Authors

Brown, HG
Ciston, J
Ophus, C

Publication Date

2019-12-19

DOI

10.1103/PhysRevResearch.1.033186

Peer reviewed

A linear-scaling algorithm for rapid computation of inelastic transitions in the presence of multiple electron scattering

Hamish G. Brown,¹ Jim Ciston,^{1,*} and Colin Ophus^{1,†}

¹*National Center for Electron Microscopy, Molecular Foundry,
Lawrence Berkeley National Laboratory, California 94720, USA[‡]*

(Dated: August 22, 2019)

Strong multiple scattering of the probe in scanning transmission electron microscopy (STEM) means image simulations are usually required for quantitative interpretation and analysis of elemental maps produced by electron energy-loss spectroscopy (EELS). These simulations require a full quantum-mechanical treatment of multiple scattering of the electron beam, both before and after a core-level inelastic transition. Current algorithms scale quadratically and can take up to a week to calculate on desktop machines even for simple crystal unit cells and do not scale well to the nano-scale heterogeneous systems that are often of interest to materials science researchers. We introduce an algorithm with linear scaling that typically results in an order of magnitude reduction in compute time for these calculations without introducing additional error and discuss approximations that further improve computational scaling for larger scale objects with modest penalties in calculation error. We demonstrate these speed-ups by calculating the atomic resolution STEM-EELS map using the L-edge transition of Fe, for of a nanoparticle 80 Å in diameter in 16 hours, a calculation that would have taken at least 80 days using a conventional multislice approach.

I. INTRODUCTION

The presence of even small numbers of dopant atoms in a material can have a disproportionate effect on that material's properties, and thus characterization techniques that perform nanoscale chemical and elemental mapping are important investigative tools in materials science. Furthermore, the presence of extended, non-periodic defects, interfaces, and nanoscale precipitate phases either naturally occurring or engineered for functionality requires quantitative atomic resolution analysis of fields of view that are substantially larger than crystal unit cells [1–4]. In scanning transmission electron microscopy electron energy-loss spectroscopy (STEM-EELS) a focused probe, which can be smaller than the width of an atom in state-of-the-art instruments, is raster scanned across a material, and the energy-loss spectrum is recorded for each scan position of the probe. Integrating regions of the spectra that correspond to energy-losses characteristic of ionization of a particular atomic element allows for mapping the positions and concentrations of elements within the specimen [5]. Additionally, analysis of the fine structure of the energy loss spectrum can reveal modified valence state of the elements due to bonding [2, 6, 7].

A complicating factor for STEM-EELS interpretation is that strong multiple scattering of the electron probe before and after exciting an ionization event means that for quantitative work, such as measuring chemical

concentrations, STEM-EELS results often need to be interpreted by performing forward simulations of an assumed structure [8]. The simulations typically combine either multislice or Bloch wave simulations with inelastic scattering cross-sections for different elements of interest [8–11]. Currently, STEM-EELS image simulations of simple crystalline structures can require up to a week of calculation time, even using graphical processing unit (GPU) accelerated simulation codes [12, 13]. Algorithms with much faster run times that scale better with system size are required to simulate large objects, or those with more heterogeneous chemical properties than simple crystals, such as samples containing dopants, extended defects, interfaces, or entire nanoparticles. Recent improvements in the readout speed of electron spectrometers adds additional impetus to efforts to speed up STEM-EELS simulations, as faster spectrometers allow for more routine imaging of much larger systems in STEM-EELS at atomic resolution.

A recent innovation in the field of electron microscope image simulation is the development of the plane wave reciprocal-space interpolated scattering matrix (PRISM) algorithm [14]. Instead of the traditional approach of calculating the propagation of the beam for each scan position in a STEM raster independently, the PRISM method uses the multislice algorithm to calculate the scattering matrix for a particular electron microscope experiment and stores it in memory. The scattering matrix, which propagates components of the electron illumination through the imaging object, can then be rapidly applied to the illumination wave function for each scan position to compute the output wave [15].

In this paper, we extend the PRISM algorithm to the problem of STEM-EELS image simulation, and

* jciston@lbl.gov

† clophus@lbl.gov

‡ hamishbrown@lbl.gov

demonstrate substantial speed up in calculation time. This paper is structured as follows: in Section II the underlying physics of STEM-EELS simulation are reviewed and we discuss both the scattering matrix operator and the multislice algorithm for numerical simulation of STEM images. In Section III we detail algorithmic implementation of both the conventional multislice approach and our novel PRISM approach to STEM-EELS calculation and discuss the predicted scaling of the run-time of both algorithms with different parameters such as simulation grid size and specimen thickness. Section IV shows results simulated using the new method, and compares accuracy and speed-up relative to the conventional multislice approach. Two further optimizations are discussed in this section: evaluating inelastic transitions on a cropped grid and using an inverse multislice operation to economize on the total number of multislice operations. These optimizations are shown to typically induce minimal errors while offering substantial reductions in compute time. This section is concluded with a calculation of the STEM-EELS image for an illustrative large heterogeneous object, in this case a FePT nanoparticle roughly 80 Å in diameter [16].

II. THEORY

In this section we briefly review the underlying physical theories behind the simulation of a relativistic electron propagating through a specimen of condensed matter. We introduce the transition potential to simulate the inelastic scattering of the probe electron due to the ionization of an electron bound to an atom within the specimen. Existing methods calculate the propagation of the electron probe to the plane of ionization and subsequent propagation of each inelastically scattered wave separately for each scan position and each inelastic transition of interest. We outline how we can calculate and store the scattering matrix operator that performs all of these steps and then rapidly apply it to all probes in the STEM raster.

We begin with the Schrödinger equation in reciprocal space for a fast electron interacting with the electrostatic potential of a specimen of condensed matter [17],

$$\frac{d\hat{\psi}_{\mathbf{g}}(z)}{dz} = -i\pi\lambda g^2 \hat{\psi}_{\mathbf{g}}(z) + \sum_{\mathbf{h}} i\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} \hat{\psi}_{\mathbf{h}}(z). \quad (1)$$

The $\hat{\psi}_{\mathbf{g}}(z)$ are Fourier coefficients of the fast electron wave function as a function of depth z in the specimen and Fourier space coordinates \mathbf{g} (magnitude g) and \mathbf{h} in the plane perpendicular to the direction of propagation. The Fourier coefficients of the electrostatic potential are denoted by $\hat{V}_{\mathbf{g}}$. The interaction constant $\sigma = 2\pi m_e e \lambda / h^2$, where m_e and λ are the (relativistically

corrected) mass and wavelength of the electron, e is the electron charge and h is Planck's constant. Equation (1) is a set of coupled linear equations for which the solution can be written as the matrix-vector product

$$\boldsymbol{\psi}(z) = \sum_{\mathbf{h}} e^{-i\pi\lambda z g^2 \delta_{\mathbf{g}\mathbf{h}} + i\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} z} \hat{\psi}_{\mathbf{h}}(0) \equiv \mathcal{S}(z) \boldsymbol{\psi}(0) \quad (2)$$

where $\delta_{\mathbf{g}\mathbf{h}}$ is the Kronecker delta and the bold $\boldsymbol{\psi}$ is a vector containing the Fourier coefficients of the illumination $\hat{\psi}_{\mathbf{h}}$. For STEM the illumination is a coherent focused probe with functional form

$$\hat{\psi}(\mathbf{g}, 0) = A(\mathbf{g}) e^{-i\chi(\mathbf{g})}. \quad (3)$$

Here $A(\mathbf{g})$ is the aperture function, a top-hat function with radius equal to the convergence semi-angle of the probe, and $\chi(\mathbf{g})$ is the aberration function which takes into account probe aberrations such as defocus, spherical aberration and astigmatism [18].

Efficient numerical calculation of Eq. (2) typically proceeds through diagonalisation of the scattering matrix \mathcal{S} (the Bloch wave method [19, 20]), or through a split-step evaluation of the action of \mathcal{S} on $\boldsymbol{\psi}$, with the specimen first split into n slices in the beam direction and the operator involving the propagation matrix elements $-\pi\lambda g^2 \delta_{\mathbf{g}\mathbf{h}} z / n_z$ and that involving the specimen interaction matrix elements $\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} z / n_z$ applied in alternating sequence:

$$\begin{aligned} \boldsymbol{\psi}(z) &= \left[e^{-i\pi\lambda z g^2 \delta_{\mathbf{g}\mathbf{h}} / n_z + i\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} z / n_z} \right]^{n_z} \boldsymbol{\psi}(0), \\ &\approx \left[e^{-i\pi\lambda z g^2 \delta_{\mathbf{g}\mathbf{h}} / n_z} e^{i\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} z / n_z} \right]^{n_z} \boldsymbol{\psi}(0). \end{aligned} \quad (4)$$

This is called the “multislice method” in the electron microscopy literature [17] and has become the most popular method of evaluation of Eq. (2). This is because the $e^{i\sigma \hat{V}_{\mathbf{g}-\mathbf{h}} z / n_z}$ operator (referred to as the transmission function) is diagonal in real space and a fast Fourier transform (an FFT, which we shall represent with the symbol $\hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{g}}$ and its inverse operation as $\hat{\mathcal{F}}_{\mathbf{g} \rightarrow \mathbf{r}}^{-1}$) [21] can be used to efficiently transform between real and reciprocal space,

$$\boldsymbol{\psi}(z) \approx \left[e^{-i\pi\lambda z / n_z g^2 \delta_{\mathbf{g}\mathbf{h}}} \hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{h}} e^{i\sigma V(\mathbf{r}) z / n_z} \hat{\mathcal{F}}_{\mathbf{g} \rightarrow \mathbf{r}}^{-1} \right]^{n_z} \boldsymbol{\psi}(0). \quad (5)$$

The two-dimensional FFT has a favourable $2N^2 \log N$ scaling, where N is the number of Fourier coefficients \mathbf{g} included in the simulation. For brevity we will define the following operator as short hand for a single multislice iteration,

$$\mathcal{M}(\Delta z) \equiv e^{-i\pi\lambda z / n_z g^2 \delta_{\mathbf{g}\mathbf{h}}} \hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{h}} e^{i\sigma V(\mathbf{r}) z / n_z} \hat{\mathcal{F}}_{\mathbf{g} \rightarrow \mathbf{r}}^{-1}. \quad (6)$$

By comparison the Bloch wave approach, which solves Eq. (2) through diagonalisation of a matrix operator

containing the coefficients within the exponent of Eq. (2), has the less favourable N^3 scaling [22]. A more in-depth explanation of the multislice algorithm with details on its implementation may be found in Ref. [18].

A recent innovation in STEM image simulation is the PRISM (plane-wave reciprocal-space interpolated scattering matrix) algorithm, in which only the rows of the matrix operator $\mathcal{S}(z)$, from Eq. (2), corresponding to the Fourier coefficients present in the illumination vector $\psi_{\mathbf{g}}$ are calculated using the multislice algorithm Eq. (5) and stored in memory. As will be explained in the next section, this approach can be used to accelerate the calculation of STEM-EELS images because the wavefunctions for the inelastically scattered electron can be rapidly propagated to the exit surface with the stored operator. This is instead of having to perform the multislice operation separately for each separate probe position and inelastic scattering event in the conventional multislice approach.

For an inelastic transition at depth z , the scattered wave function $\psi_n(\mathbf{r}, z)$, where n is shorthand for an excited quantum state, is given by the product of the elastic wavefunction $\psi_0(\mathbf{r}, z)$ and an inelastic transition potential for ionization $H_{n0}(\mathbf{r}, z)$,

$$\psi_n(\mathbf{r}_{\perp}, z) = H_{n0}(\mathbf{r}_{\perp})\psi_0(\mathbf{r}_{\perp}, z). \quad (7)$$

The inelastic transition potential is given by Ref. [23],

$$H_{n0}(\mathbf{r}) = \int u_n^*(\boldsymbol{\kappa}, \mathbf{r}') \frac{e^2}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|} u_0(\mathbf{r}') d\mathbf{r}' \quad (8)$$

where $u_0(\mathbf{r}')$ is the wave function of an electron in a bound state of the specimen and $u_n(\boldsymbol{\kappa}, \mathbf{r}')$ is the wavefunction of that electron excited to a continuum (free) state with wave vector $\boldsymbol{\kappa}$. The wave function in the continuum state is typically expanded in a spherical harmonic basis,

$$u_n(\boldsymbol{\kappa}, \mathbf{r}) = \frac{4\pi}{2\kappa r} \sum_{\ell'=0}^{\infty} i^{\ell'} e^{i\delta_{\ell'}} u_{\kappa\ell'}(r) \times \sum_{m_{\ell'}=-\ell'}^{\ell'} Y_{\ell', m_{\ell'}}^*(\hat{\boldsymbol{\kappa}}) Y_{\ell', m_{\ell'}}(\hat{\mathbf{r}}) \quad (9)$$

where the $Y_{\ell', m_{\ell'}}$ are spherical harmonic functions, ℓ , m_{ℓ} , ℓ' and $m_{\ell'}$ are the angular momentum and azimuthal angular momentum quantum numbers of the bound and continuum states and $\delta_{\ell'}$ is a phase factor determined by fitting to an asymptotic form of the free state [23, 24]. To a good approximation, only a small number of the different possible $(\ell', m_{\ell'})$ need be included in a converged calculation [13]. The kernel in Eq. (8) is the Coulomb potential, which mediates interactions between the fast electron and the sample, and ϵ_0 is the permittivity of free space. Calculations of Eq. (8) used in this paper are based on a solution using an angular

momenta basis for u_0 and u_n which is derived in Ref. [23], the numerical implementation of which is discussed in Ref. [24].

We may simulate a STEM-EELS image by then propagating the wave function for the inelastically scattered electron, ψ_n from Eq. (8), to the exit surface using Eq. (2). We represent this process mathematically

$$\psi_n(\mathbf{r}, t) = \mathcal{S}_1 \mathcal{F} \mathcal{H}_{n0} \mathcal{F}^{-1} \mathcal{S}_2 \psi_0(0) \quad (10)$$

$$\equiv \mathcal{S}_n \psi_0(0). \quad (11)$$

Here \mathcal{H}_{n0} is a matrix containing the values of $H_{n0}(\mathbf{r})$ on its diagonals and the matrix operator \mathcal{S}_n we have defined in Eq. (11) for imaging of a single inelastic transition. Equation (10) must be solved for each inelastic transition of interest. Figure 1 is a diagrammatic representation of Eq. (11), Fig. 1(a) shows how \mathcal{S}_1 maps Fourier coefficients of the probe wave function to wave functions at the depth of the inelastic transition (b) where we multiply these wave functions by the ionization transition potentials in Fig. 1(c) and then propagate them using \mathcal{S}_2 , Fig. 1(d), to points within the EELS aperture in Fig. 1(e).

III. ALGORITHMS

In this section we outline the details of the implementation of the simulation algorithms and write down estimates for the runtimes of each algorithm. We start with the algorithm for calculating STEM-EELS images using the multislice method:

```

Loop over probe positions
for  $x=1$  to  $n_x$  do
  for  $y=1$  to  $n_y$  do
    Initialise illumination wave function
     $\psi(\mathbf{g}, 0) = A(\mathbf{g})e^{-2\pi i(x,y) \cdot \mathbf{k}}$ 
    Loop over slices of specimen
    for  $i_z = 1$  to  $n_z - 1$  do
      Loop over inelastic transitions within slice  $i_z$ 
      for  $n = 1$  to  $n_{\text{states}, i}$  do
        Calculate inelastically scattered electron
        wavefunction  $\psi_n(\mathbf{r}_{\perp}, i_z) = H_{n0}(\mathbf{r}_{\perp})\psi_0(\mathbf{r}_{\perp}, i_z)$ 
        Propagate  $\psi_n$  to exit surface of specimen
        for  $i'_z = i_z$  to  $n_z$  do
          Multislice to advance  $\psi_n$  one slice
           $\psi_n(\mathbf{g}, i'_z) = \mathcal{M}(\Delta z)\psi_n(\mathbf{h}, i'_z - 1)$ 
        end for
        Integrate wavefunction over detector function
         $D(\mathbf{g})$  and add contribution to STEM image  $I(x, y) =$ 
 $I(x, y) + \int |\psi_n(\mathbf{g}, n_z)|^2 D(\mathbf{g}) d\mathbf{g}$ 
      end for
      Multislice to advance  $\psi_0$  one slice
       $\psi(\mathbf{g}, i_z + 1) = \mathcal{M}(\Delta z)\psi(\mathbf{h}, i_z)$ 
    end for
  end for
end for

```

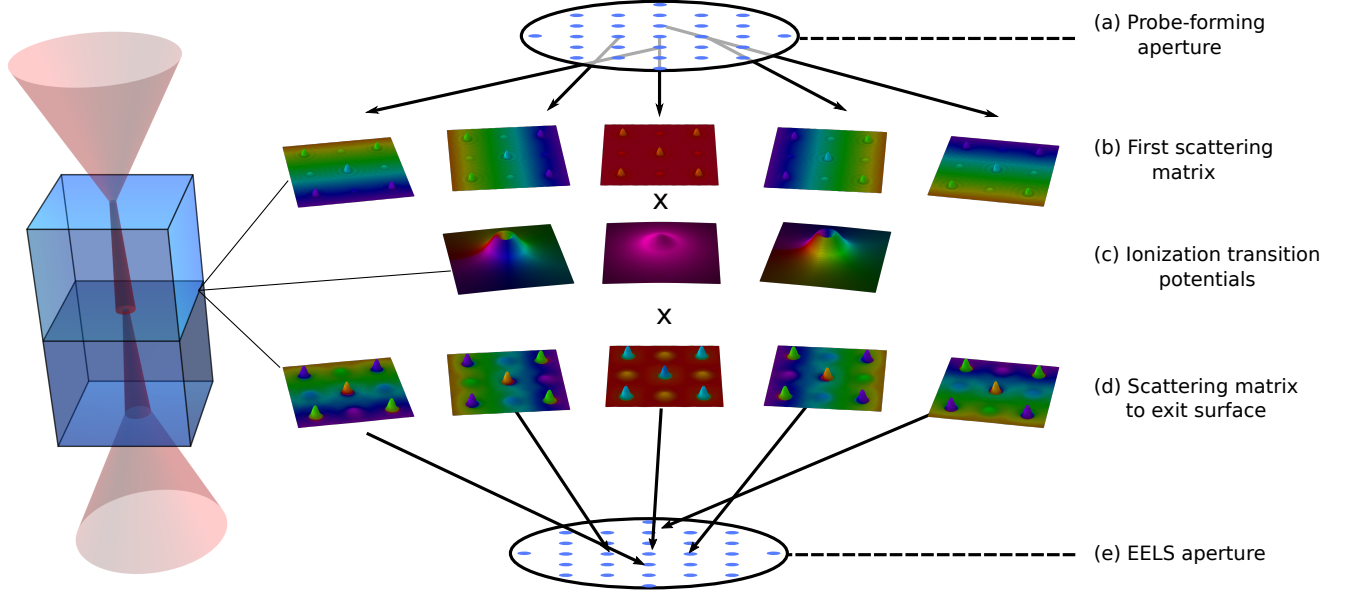


FIG. 1. A diagrammatic representation of the PRISM algorithm for STEM-EELS simulation that is encapsulated in Eq. (10). The first scattering matrix maps points in the probe-forming aperture (a) to complex output in the plane of the ionization (b). At this plane the output is multiplied by the ionization transition potentials of interest H_{n0} which are shown in (c) and then by the output of the second scattering matrix (d) which propagates the resulting inelastically scattered wave to reciprocal space points inside the EELS aperture (e).

Assuming a constant number of inelastic transitions n_{states} at each slice, which is not true in general but is a useful assumption for estimating the computational complexity of the above algorithm, this algorithm requires

$$N_{MS} = \sum_{i_z=0}^{n_z-1} [n_{states}(n_z - i_z)] + n_z \quad (12)$$

$$= n_{states}n_z(n_z - 1)/2 + n_z \quad (13)$$

multislice operations for each probe position. The arithmetic sum identity $\sum_{i=0}^{n-1} i = n(n-1)/2$ has been invoked in the above equation.

In our proposed PRISM STEM-EELS algorithm we calculate two scattering matrices at each slice, the first to propagate the probe wave function to the slice of the inelastic transition and the second to propagate the wave function for the inelastically scattered electron to the exit surface.

Initialize scattering matrices

$$\mathcal{S}_1 = \mathcal{I}$$

$$\mathcal{S}_2 = [\mathcal{M}^T(\Delta z)]^{n_z}$$

for $i_z = 1$ to n_z **do**

 Loop through transitions within slice i_z

for $n = 1$ to n_{states, i_z} **do**

 Calculate Eq. (11)

$$\mathcal{S}_{n, i_z} = \mathcal{S}_2 \mathcal{F} \mathcal{H}_{n0} \mathcal{F}^{-1} \mathcal{S}_1$$

Apply \mathcal{S}_{n, i_z} to each illumination vector in the raster scan and add the resulting amplitude to the STEM image

$$I(x, y) = I(x, y) + |\mathcal{S}_{n, i_z} \psi(\mathbf{g}, x, y)|^2$$

end for

Advance \mathcal{S}_1 one slice $\mathcal{S}_1(i_{z+1}) = \mathcal{M}(\Delta z) \mathcal{S}_1(i_z)$

Re-calculate \mathcal{S}_2 , $\mathcal{S}_2 = [\mathcal{M}^T(\Delta z)]^{n_z - i_z} \mathcal{I}$

end for

An important point to note is that, as defined here, the rows of \mathcal{S}_1 correspond to points in reciprocal space within the probe forming aperture and the columns correspond to real space points in the specimen plane, so \mathcal{S}_1 is generally a non-square matrix due to the different sampling of these two planes. Each column can be calculated by using multislice to propagate the plane wave components that fall within the probe forming aperture through the specimen, as shown in Fig. 1 [14]. Conversely, the rows of \mathcal{S}_2 correspond to real space points in the specimen plane and the columns correspond to points within the EELS detector. The most efficient way to calculate \mathcal{S}_2 is to propagate plane wave components with transverse momenta within the EELS acceptance aperture *back* through the specimen. This is formally the transpose of the multislice operation defined in Eq. (6), which is represented by a superscript T in the above algorithm description, and is given by

$$\mathcal{M}^T(\Delta z) \equiv \hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{h}} e^{i\sigma V(\mathbf{r})\Delta z} \hat{\mathcal{F}}_{\mathbf{g} \rightarrow \mathbf{r}}^{-1} e^{-i\pi\lambda\Delta z g^2 \delta_{gh}}. \quad (14)$$

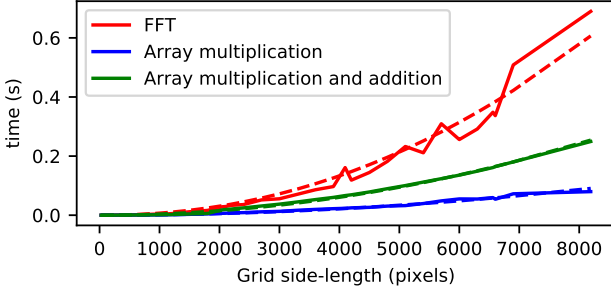


FIG. 2. Scaling of FFT, array multiplication and array multiplication and addition operations in Matlab with the side length (pixels) of the square grid. The fitted scalings for these operations are plotted with dashed lines and the coefficients of the fit are given in the text.

The PRISM STEM-EELS algorithm requires $n_z * [N_1 + N_2(n_z + 1)]$ multislice operations where N_1 is the number of rows in matrix \mathcal{S}_1 , which correspond reciprocal space points in the illumination, and N_2 is the number of columns in matrix \mathcal{S}_2 , which correspond to reciprocal space points in the EELS detector. Since the PRISM STEM-EELS simulation algorithm is typically more economical with the required number of multislice iterations, the matrix multiplication step, evaluating Eq. (10), tends to be the rate limiting step. This means that although the calculation time is technically quadratic in n_z , in many cases run time scaling is instead predominantly determined by the number of unique transitions. If we further assume that the number of transitions in each slice is roughly constant then the scaling will be roughly linear with the number of slices n_z . In the upcoming Sec. IV C, we will introduce an approximation that makes the scaling truly linear with n_z .

Comparing the run time of the PRISM STEM-EELS algorithm to the conventional multislice approach requires accounting for the relative speed of FFTs to array multiplication. To explore this question we make the approximation that the computation time of a two-dimensional FFT can be parametrized as $T_{\text{FFT}} = AN^2 \log N$ and that of an array multiplication (as used in the multislice operation) can be parametrized as $T_{\text{mult}} = BN^2$ and that of an array multiplication and summation step (as would occur in a step of a matrix multiplication where a single row and column are multiplied and added to the final result) can be parametrized $T_{\text{addmult}} = CN^2$. Here A , B and C are constants and N is the size of the (square) grid using in computation. Measurements of T_{FFT} , T_{mult} and T_{addmult} for different values of N for Matlab are plotted in Fig. 2 for which values of $A = 1.0 \times 10^{-9}$, $B = 9.0 \times 10^{-9}$ and $C = 3.8 \times 10^{-9}$ were fitted (fitted functions are plotted with dashed lines). However, for the most accurate estimates of the runtime of calculations at a specific pixel grid size N , in the

examples provided in this paper we will simply measure T_{FFT} , T_{mult} and T_{addmult} at that pixel grid size N . We also note that the platform and compiler can impact the relative speeds of different algorithms.

We assume a simulation object measuring $L \times L$ in the plane perpendicular to beam propagation and of thickness z , which we divide into $n_z = z/\Delta z$ slice. This object would require a STEM scan with a Nyquist sampling of $(4L\alpha)^2$ probe positions, where α is the probe forming aperture in units of inverse length [12]. Assuming a constant number of ionization states at each slice, not true in general as atoms of the element of interest might be not be uniformly dispersed throughout the sample, but a useful approximation for the current timing estimates, the computation time for the conventional multislice approach will be given by

$$T_{MS} \approx 2(4L\alpha)^2(n_{\text{states}}n_z(n_z - 1)/2 + n_z) (AN^2 \log N + BN^2). \quad (15)$$

The speed of the PRISM STEM-EELS calculation will depend on the size of the matrices used. For matrix \mathcal{S}_1 , the number of rows will be depend on the sampling of the probe forming aperture function $A(\mathbf{q})$ in Eq. (3) which covers a reciprocal space area of $\pi\alpha^2 \text{ \AA}^{-2}$. A simulation cell size of L implies a natural reciprocal space sampling of L pixels per unit of inverse length for the illumination [18] though this can be further reduced to L/f by using a PRISM interpolation factor of f – an optimization described in detail in Ref. [14]. The number of rows in the scattering matrix will then be given by $\pi\alpha^2 L^2/f^2$. By similar reasoning, the second scattering \mathcal{S}_2 matrix, which propagates the inelastically scattered electrons from the plane of ionization to the EELS aperture, will have a number of columns equal to $\pi\beta^2 L^2/f^2$ where β is the diffraction space size of the EELS aperture. The time required for multislice iterations necessary for the PRISM algorithm is therefore given by

$$T_{\text{PRISM, multislice}} = 2(\pi\alpha^2 + 2 * (n_z - 1)\pi\beta^2)L^2/f^2 \cdot n_z(AN^2 \log N + BN^2). \quad (16)$$

The number of columns in \mathcal{S}_1 and the number of rows in \mathcal{S}_2 will be given by the square of half total number of pixels in the simulation cell $(N/2)^2$. These matrices need only have a side-length of $N/2$ since the output from a multislice calculation is bandwidth limited either to 1/2 (as in the implementation used for this investigation) or 2/3 of the total array size and only spatial frequencies within this band-limit need be kept in the scattering matrix. For more detail on the need for this bandwidth limiting approach the reader is referred to Sec. 6.8 of Ref. [18]. With reference to the STEM-EELS simulation using scattering matrices we must sum the computation times of the multislice iterations and

the matrix multiplications,

$$T_{\text{PRISM}} \approx B(N/2)^2 \pi \alpha^2 L^2 / f^2 + C(N/2)^2 \pi^2 \alpha^2 \beta^2 L^4 / f^6 + 4C\pi^2 \alpha^4 \beta^2 L^6 / f^6 \quad (17)$$

Where the first term is the time required to calculate the matrix multiplication of $\mathcal{S}_1 \mathcal{H}_{n0}$ in Eq. (10), the second term the time required to do the second matrix multiplication ($\mathcal{S}_1 \mathcal{H}_{n0}$) \mathcal{S}_2 in Eq. (10) and the third term the time required to do the matrix multiplication $\mathcal{S}_n \Psi$ for each probe ψ in the STEM raster in Eq. (11). In the following section we discuss a MATLAB implementation of the STEM-EELS simulation algorithms discussed show that the above expressions, Eqs. (15), (16) and (17), are correct estimators of the runtime of these calculations.

Memory requirements for both algorithms also merit discussion. At a minimum the multislice algorithm requires only arrays containing the Fresnel free-space propagator, transmission function, probe and ionization transition potential, so 4 $N \times N$ complex valued arrays. If there is sufficient memory, as is the case for most simulations, the transmission functions for all the slices will also be stored rather than calculated on the fly – an additional $n_z N \times N$ complex arrays. The PRISM STEM-EELS algorithm adds the requirement that two scattering matrices be stored in memory, an additional $\pi \alpha^2 / f^2 + \pi \beta^2 / f^2$ arrays of size $N/2 \times N/2$. By way of example we discuss the requirements for simulation of the nanoparticle performed in the upcoming Sec. IV D. The object was sampled on a 1836×1836 grid and partitioned depth wise into 45 slices. The minimum of 4 single-precision complex floating-point arrays for the conventional multislice approach would take up 107 MB of memory. Storing the transmission functions takes up a further 1.21 GB. The number of columns in the first scattering matrix \mathcal{S}_1 and the number of rows in the second scattering matrix \mathcal{S}_2 are both 325. Each of these S-matrix rows and columns has size equal to a 918×918 grid so approximately 4.38 GB is required to store both matrices. This is a substantially greater amount of memory than is the case for the equivalent multislice calculation, though we note that with current technology most high-end graphics cards have 8 GB or greater of memory and this example calculation is for a larger simulation cell than has typically been attempted before.

IV. RESULTS AND DISCUSSION

A. Implementation of a scattering matrix based STEM-EELS simulation method

In this section we report results from a Matlab implementation of the conventional multislice algorithm

and the new PRISM algorithm. This implementation is included in the supplementary materials of this paper. It is included here to provide an accessible demonstration of the two algorithms introduced in the previous Sec. II; the code does not calculate the inelastic transition potentials H_{n0} given by Eq. (10) but uses a single calculated H_{n0} outputted from the μSTEM code [10]. As a test case we calculate STEM-EELS images of a single transition ($\ell = 0, m_\ell = 0 \rightarrow \ell' = 1, m'_\ell = 1$) for ionization of the O 1s orbital (the K-edge) using both the conventional multislice method (red solid line) and the new PRISM approach (blue solid line) for thicknesses between 10 Å and 100 Å. A 2×2 tiling of the SrTiO₃ unit cell (measuring $7.81 \text{ Å} \times 7.81 \text{ Å}$) specimen and a 160×160 pixel grid was used, parameters which are likely to result in somewhat unconverged calculations but result in faster runtimes for both algorithms and so allow rapid comparison of results. For the purposes of simplifying comparison of conventional multislice and PRISM-EELS results thermal vibrations of the atoms were turned off for this calculation. The actual timings for each algorithm are compared in Fig. 3 with the relevant estimates from Eq. (15) for the multislice case using measured T_{FFT} , T_{mult} and T_{addmult} for a 160×160 pixel grid (red dashed line) and the sum of Eqs. (16) and (17) for the PRISM case (blue dashed line) showing that these equations give reasonable estimates of computation time for these simulations. Images for the PRISM and conventional multislice calculations are shown for the thicknesses 10, 50 and 80 Å. We compare differences in the images using the following normalized root mean sum of squares error percentage error metric,

$$\varepsilon = \frac{1}{100} \sqrt{\frac{(\sum I_{\text{MS}}(R) - I_{\text{PRISM}})^2}{(\sum I_{\text{MS}}(R))^2}}. \quad (18)$$

These are tabulated in Fig. 4, both for the total image (Total error, ε_T) and in a 6×6 pixel window centered on the Ti-O column (site error ε_S). Total error ε_T is typically less than 0.01% and site error ε_S is between 0.001% and 0.02%. This low discrepancy confirms that both the PRISM and conventional multislice approaches indeed encapsulate the same scattering physics. The PRISM method is faster for all thicknesses and exhibits the pseudo-linear scaling predicted for this algorithm, whilst the scaling of the conventional multislice method is quadratic with thickness.

B. Algorithm speed ups: calculating inelastic transitions on a cropped grid.

Further speed-ups, with a modest penalty to accuracy, can be achieved by only evaluating the matrix multiplication in Eq. (11) in a fraction of the grid centered about the site location of the transition event $H_{n0}(\mathbf{r})$ rather than over the whole simulation grid. To

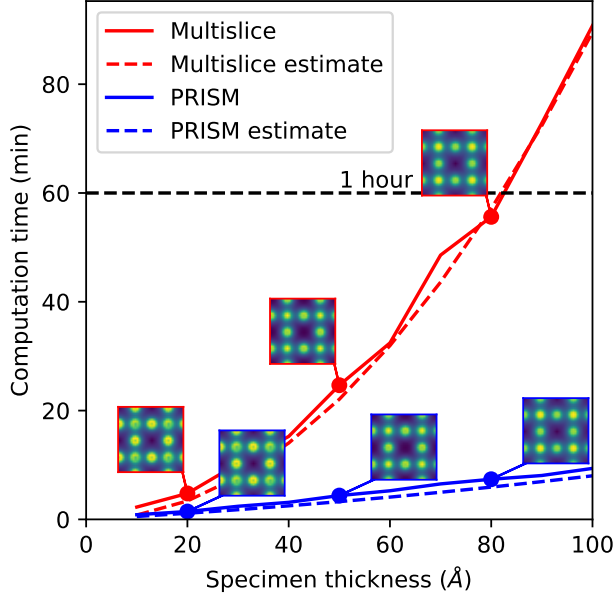


FIG. 3. Scaling of computation time for standard multislice and PRISM approaches to the STEM-EELS image simulation with thickness for a SrTiO₃ sample, calculation details are given in the text.

Thickness (Å)	Total error (ε_T)	Site error (ε_S)
20	0.00207 %	0.00186 %
50	0.00654 %	0.00855 %
80	0.00779 %	0.00783 %
100	0.01040 %	0.01580 %

FIG. 4. Percentage error of PRISM calculation relative to multislice calculation for each thickness in Fig. (3)

demonstrate this optimization we simulate STEM-EELS images of a single Mn dopant occupying an Sr site midway through a 50 Å thick SrTiO₃ crystal. An L-shell transition for Mn is shown in Fig. 5 where Eq. (11) is (a) evaluated over the full grid (which has a side length of approximately 8 Å), and in cropped regions with side-lengths measuring approximately (b) 4 Å and (c) 2 Å. The size of these windows relative to the transition potential $H_{n0}(\mathbf{r})$ is indicated in Fig. 5(d). Only the transitions ($\ell = 1, m_\ell = 1$) \rightarrow ($\ell' = 2, m_{\ell'} = 2$) and ($\ell = 1, m_\ell = -1$) \rightarrow ($\ell' = 2, m_{\ell'} = -2$) for energy losses 1 eV over the ionization threshold where included for the purpose of this demonstration. The percentage errors, relative to the image in Fig. 5(a), are written on images Fig. 5(b) and (c). Both the total image error ε_T and error for the atomic site ε_S , which is evaluated only for a 1 Å window around the Mn atomic position are indicated. The site error ε_S is the most relevant metric for our purposes since elemental concentration mapping would typically proceed by integrating the STEM-EELS signal in a window centered on the atomic site and then relating the result to a pre-computed look-up table. A

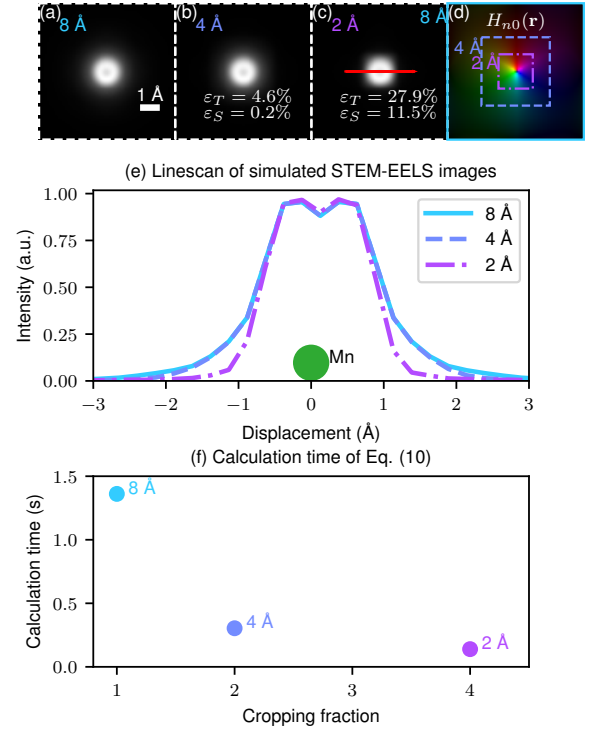


FIG. 5. The STEM-EELS image for a single L shell transition for a dopant Mn atom within a 100 Å thick SrTiO₃ crystal is calculated using different grids of side length (a) 8 Å (b) 4 Å and (c) 2 Å to evaluate Eq. (11). The size of these windows is shown relative to the complex $H_{n0}(\mathbf{r})$ transition potential in the rightmost panel of (d). (e) A linescan through the image, indicated with a red arrow in (c), shows differences in the tails of the image. (f) The time required to calculate Eq. (11) for each of the images in (a)-(c).

site error of 0.2%, as in Fig. 5(b) is likely acceptable, though a site error of 11.4% as in Fig. 5(c) is likely too high, suggesting the cropping window chosen was too aggressive. Inspection of a linescan in Fig. 5(e) shows that in the image from Fig. 5(c) the long tails of the Mn transition potential H_{n0} have been cropped out. Figure 5(f) details the significant speed up benefits of this approach. The speed ups are quadratic which is consistent with the array multiplication scaling quadratically with grid pixel size, as the calculation time of Eq. (11) is observed to roughly quarter with each halving of the window in Fig. 5(a). For this particular transition potential a cropping box with side length of 4 Å gives the best balance between calculation time speed up and loss of accuracy.

Thickness (Å)	Total error (ε_T)	Site error (ε_S)	Speed up (%)
20	0.17 %	0.31 %	2.61 %
50	0.72 %	1.28 %	16.0 %
80	0.97 %	1.56 %	24.4 %
100	1.00 %	1.53 %	21.3 %

FIG. 6. Percentage errors in the PRISM STEM-EELS calculations with the inverse multislice optimization relative to conventional multislice calculations as well as the percentage speed up relative to PRISM STEM-EELS calculations without the inverse multislice optimization.

C. Algorithm speed ups: inverse multislice

In the PRISM STEM-EELS algorithm described in Sec. IV A, the scattering matrix \mathcal{S}_2 which propagates the inelastically scattered electron wave ψ_n from the depth at which ionization occurred to the exit surface is calculated from scratch for each thickness i_z . This is a duplication of work, since \mathcal{S}_2 was at some point calculated for all thicknesses in the initialization step $\mathcal{S}_2 = (\mathcal{M})^T n_z$. One approach would be to store \mathcal{S}_2 for each slice, which is not practical for most calculations given the size of the scattering matrix (a complex numbered array of size $\pi\alpha^2 L^2 / f^2 \times N \times N$). A second approach which sacrifices some accuracy at the expense of calculation time would be to perform the inverse of the multislice operation (\mathcal{M}^{-1}) to retreat the scattering matrix \mathcal{S}_2 a single slice. The inverse multislice operation is defined

$$\mathcal{M}^{-1}(\Delta z) \equiv \hat{\mathcal{F}}_{\mathbf{h} \rightarrow \mathbf{r}} e^{-i\sigma V(\mathbf{r})\Delta z} \hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{g}} e^{i\pi\lambda\Delta z g^2 \delta_{gh}}. \quad (19)$$

Relative to the forward multislice operation defined in Eq. (6) the order of FFT, multiplication, inverse FFT and propagation steps has been reversed and the complex-conjugate of the transmission function and propagation operators is used instead. This can introduce some error since a forward multislice iteration can cause electrons to scatter to high angles outside the bandwidth limit of the calculation and these electrons will not be recovered with an inverse operation. For the simulation of the SrTiO_3 in Sec. IV A we report the percentage errors, both total error ε_T and site error ε_S , in Fig. 6. For this case, the errors relative to conventional multislice calculations are found to be around 1%, which is small whilst the speed up is around 20% relative to PRISM calculations for some of the thicker cases considered without the inverse multislice optimization.

D. Calculations for heterogeneous nanometer scale objects

In Sec. IV A we showed that using the PRISM approach for STEM-EELS simulations resulted in computation times that scale more favourably with specimen thickness

and that calculation time can be further reduced by evaluating the inelastic scattering cross section, Eq. (11), only in a fraction of the calculation grid centered on the transition of interest and using the inverse multislice optimization. In this section we demonstrate how these improvements, taken together, allow calculations of much larger objects than has previously been feasible, in this case a Fe-Pt nanoparticle that is approximately 80 Å in diameter. The atomic coordinates for this nanoparticle were reconstructed from a STEM-HAADF tomography tilt series in Ref. [16] and the object contains 6,569 Fe atoms and 16,627 Pt atoms. The projected potential of the nanoparticle is plotted in Fig. 7(a) with Pt potential shown in blue and Fe atoms shown red. The potential resulting from only the Fe atoms is plotted separately in Fig. 7(b). Both potentials have been convolved with a Gaussian function ($\sigma = 2$ pixels) to make viewing easier.

We consider the $(\ell = 1, m_\ell = 0) \rightarrow (\ell' = 2, m'_\ell = 1)$, $(\ell = 1, m_\ell = 0) \rightarrow (\ell' = 2, m'_\ell = -1)$, $(\ell = 1, m_\ell = 1) \rightarrow (\ell' = 2, m'_\ell = 2)$ and $(\ell = 1, m_\ell = -1) \rightarrow (\ell' = 2, m'_\ell = -2)$ transitions which between them account for just over 90% of total transitions for energy losses 1 eV over the ionization threshold Fe L edge. For the PRISM-EELS calculation, we use a PRISM interpolation factor of 9, such that each individual probe will be effectively calculated on a $10 \text{ Å} \times 10 \text{ Å}$ grid, a probe step of 0.246 Å (Nyquist sampling) and evaluate the inelastic scattering cross section [Eq. (11)] on a $4 \text{ Å} \times 4 \text{ Å}$ grid (which was seen to give the best trade-off between calculation speed and accuracy in Fig. 5). Then, Eq. (16) and Eq. (17) estimate the run-time of such a calculation to be 2 days. The results of this PRISM STEM-EELS simulation, which in reality took 16 hours, are shown in Fig. 7(c). We also estimate the computation time of a conventional multislice simulation using Eq. (15) for the full nanoparticle with the same probe step size and 2 Å slices along the beam direction. If further approximations are made to use only 1/9 of the grid for probe propagation and only evaluating transitions within 4 Å of the probe (i.e. those that we deem to have a reasonable chance of being excited), multislice simulation could be run in about 87 days for a single frozen phonon pass [26].

The simulated image in Fig. 7(c) is qualitatively similar to the Fe potential Fig. 7(b) but artifacts due to strong scattering of the beam are evident. In particular, the center of the nanoparticle has a lower intensity than would be expected from the density of Fe, due to the strong high-angle scattering of the electron beam from the heavier Pt atoms. Close inspection of many of the Fe columns in this region reveals small regions of lower intensity, giving rise to a “donut” or “volcano” structure which is evident in the zoomed region shown in the bottom left hand corner of Fig. 7(c). These features result from inelastic scattering of beam electrons by the heavier Pt atoms, and to a smaller extent by the Fe atoms, to high angles outside the acceptance angle of

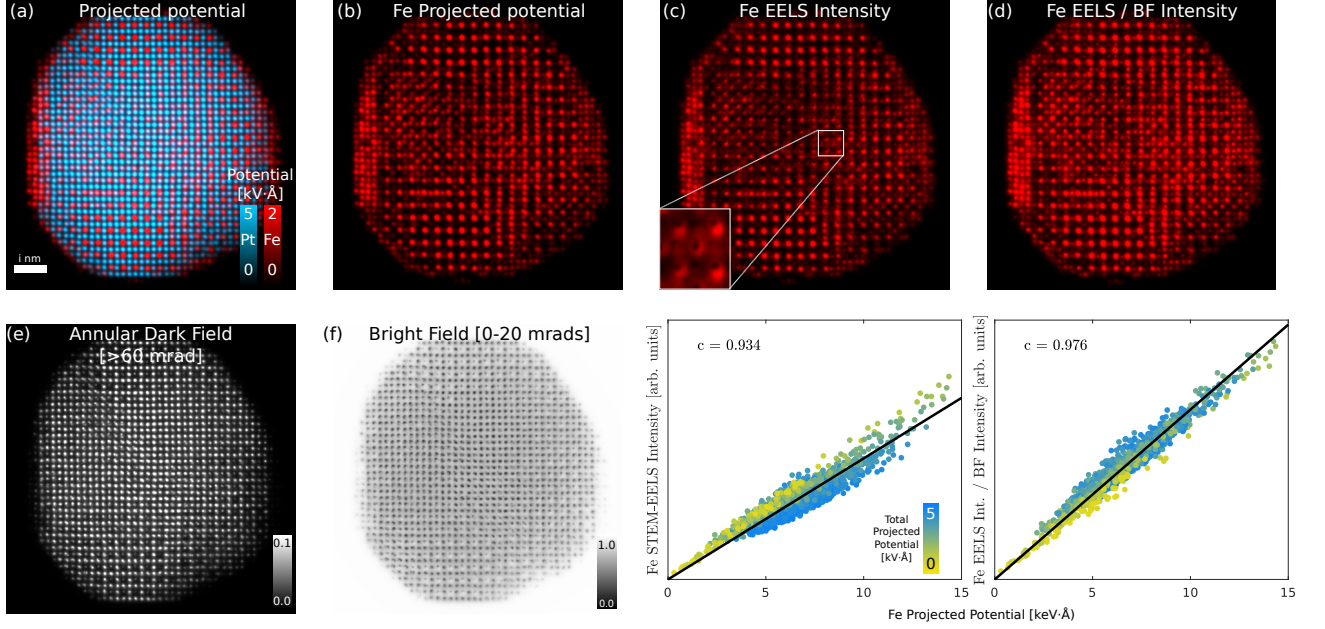


FIG. 7. (a) The projected electrostatic potential for the FePt nanoparticle reconstructed in Ref. [16] with Pt atoms indicated in blue and Fe atoms in red, the Fe projected potential is shown separately in (b). The STEM-EELS simulation of the structure in (c), which is corrected by dividing by an incoherent bright-field image in (d) as suggested in Ref. [25]. The annular dark field and incoherent bright-field images are shown in (e) and (f). Figures (g) and (h) plot the relationship between Fe Projected potential and STEM-EELS intensity, averaged in 5×5 pixel windows, for (c) and (d) respectively.

the EELS detector when the beam is scanned atop an atomic column. Even if these scattered electrons do cause the ionization of an Fe atom they are unlikely to finally contribute toward the final STEM-EELS image. A detailed explanation of this phenomenon can be found in Ref. [25] along with a strategy to obtain an image more amenable to direct interpretation: dividing by a simultaneously recorded STEM incoherent bright field image with detector of equal angular extent to the EELS aperture. This image is displayed in Fig. 7(d) which is indeed observed to be a more faithful representation of the projected electrostatic potential due to Fe in Fig. 7(b). The STEM annular dark field image, formed with a detector of inner angle 60 mrad and the incoherent bright-field image are shown in Figs. 7(e) and (f) for reference.

STEM-EELS simulations of objects of this size allow statistical analysis of STEM-EELS images as shown in Fig. 7(g), which plots the integrated potential just for the Fe atoms (a proxy for the projected Fe density) against the STEM-EELS intensity in for each 5×5 pixel region in the image. The colour of each point in the scatter plot corresponds to the total projected potential from *both* Fe and Pt with reference to the colorbar in the figure. The relationship between STEM-EELS intensity and projected Fe density is approximately linear. However there are noticeable systematic deviations with points falling below the trendline tending to be those with a higher total projected potential – a clear demonstration

of how strong elastic and inelastic scattering of the beam complicates direct interpretation of the STEM-EELS maps. These systematic errors are mostly remedied by the division of the incoherent bright-field image, which gives an improvement in the Pearson correlation score from 0.935 to 0.976. This correlation value measures the quality of the fitted trendline, showing that there is less systematic deviation from the linear relationship of Fe density and EELS intensity after applying the BF correction step. This insight demonstrates how our faster STEM-EELS algorithm, by virtue of its better scalability to larger simulation grids, can give a valuable insights into interpretation of heterogenous nanoscale STEM-EELS maps.

V. CONCLUSION

We have developed a new algorithm for simulating STEM-EELS results that economises on the number of multislice iterations required. This algorithm should run faster in general as the calculation time typically scales linearly with specimen thickness, whilst the conventional algorithm scales quadratically with specimen thickness. We have shown that with a some penalty to accuracy, even faster calculation times are possible. Finally, we have also shown that our algorithm can be used to simulate larger nanoscale objects than was previously the

case.

VI. ACKNOWLEDGMENTS

The authors thank Les Allen and Scott Findlay for helpful discussions and proofreading during the preparation of this manuscript. Work at the Molecular

Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. All authors acknowledge additional support from the U.S. Department of Energy Early Career Research Program.

VII. BIBLIOGRAPHY

-
- [1] H. L. Xin, J. A. Mundy, Z. Liu, R. Cabezas, R. Hovden, L. F. Kourkoutis, J. Zhang, N. P. Subramanian, R. Makharia, F. T. Wagner, *et al.*, Atomic-resolution spectroscopic imaging of ensembles of nanocatalyst particles across the life of a fuel cell, *Nano letters* **12**, 490 (2011).
 - [2] J. A. Mundy, Q. Mao, C. M. Brooks, D. G. Schlom, and D. A. Muller, Atomic-resolution chemical imaging of oxygen local bonding environments by electron energy loss spectroscopy, *Applied Physics Letters* **101**, 042907 (2012).
 - [3] J. A. Mundy, Y. Hikita, T. Hidaka, T. Yajima, T. Higuchi, H. Y. Hwang, D. A. Muller, and L. F. Kourkoutis, Visualizing the interfacial evolution from charge compensation to metallic screening across the manganite metal-insulator transition, *Nature communications* **5**, 3464 (2014).
 - [4] R. R. Zamani, F. S. Hage, S. Lehmann, Q. M. Ramasse, and K. A. Dick, Atomic-resolution spectrum imaging of semiconductor nanowires, *Nano letters* **18**, 1557 (2017).
 - [5] R. F. Egerton, *Electron energy-loss spectroscopy in the electron microscope* (Springer Science & Business Media, 2011).
 - [6] D. A. Muller, Y. Tzou, R. Raj, and J. Silcox, Mapping sp^2 and sp^3 states of carbon at sub-nanometre spatial resolution, *Nature* **366**, 725 (1993).
 - [7] D. A. Muller, L. F. Kourkoutis, M. Murfitt, J. Song, H. Hwang, J. Silcox, N. Dellby, and O. Krivanek, Atomic-scale chemical imaging of composition and bonding by aberration-corrected microscopy, *Science* **319**, 1073 (2008).
 - [8] M. Bosman, V. J. Keast, J. L. Garcia-Munoz, A. J. D'Alfonso, S. D. Findlay, and L. J. Allen, Two-dimensional mapping of chemical information at atomic resolution, *Physical Review Letters* **99**, 086102 (2007).
 - [9] L. J. Allen, D. C. Bell, T. W. Josefsson, A. E. C. Spargo, and S. L. Dudarev, Inner-shell ionization cross sections and aperture size in electron energy-loss spectroscopy, *Physical Review B* **56**, 9 (1997).
 - [10] L. J. Allen, S. D. Findlay, *et al.*, Modelling the inelastic scattering of fast electrons, *Ultramicroscopy* **151**, 11 (2015).
 - [11] I. Lobato, S. Van Aert, and J. Verbeeck, Progress and new advances in simulating electron microscopy datasets using MULTTEM, *Ultramicroscopy* **168**, 17 (2016).
 - [12] C. Dwyer, Simulation of scanning transmission electron microscope images on desktop computers, *Ultramicroscopy* **110**, 195 (2010).
 - [13] C. Dwyer, The role of symmetry in the theory of inelastic high-energy electron scattering and its application to atomic-resolution core-loss imaging, *Ultramicroscopy* **151**, 68 (2015).
 - [14] C. Ophus, A fast image simulation algorithm for scanning transmission electron microscopy, *Advanced structural and chemical imaging* **3**, 13 (2017).
 - [15] A. Pryor, C. Ophus, and J. Miao, A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy, *Advanced structural and chemical imaging* **3**, 15 (2017).
 - [16] Y. Yang, C.-C. Chen, M. C. Scott, C. Ophus, R. Xu, A. Pryor, L. Wu, F. Sun, W. Theis, J. Zhou, *et al.*, Deciphering chemical order/disorder and material properties at the single-atom level, *Nature* **542**, 75 (2017).
 - [17] D. Van Dyck, Image calculations in high-resolution electron microscopy: problems, progress, and prospects, in *Advances in electronics and electron physics*, Vol. 65 (Elsevier, 1985) pp. 295–355.
 - [18] E. J. Kirkland, *Advanced computing in electron microscopy* (Springer Science & Business Media, 2010).
 - [19] H. Bethe, Theorie der beugung von elektronen an kristallen, *Annalen der Physik* **392**, 55 (1928).
 - [20] C. J. Humphreys, The scattering of fast electrons by crystals, *Reports on Progress in Physics* **42**, 1826 (1979).
 - [21] We use the following convention, a forward Fourier transform is given by $\hat{\psi}(\mathbf{g}) = \hat{\mathcal{F}}_{\mathbf{r} \rightarrow \mathbf{g}} \psi(\mathbf{r}) = \int \psi(\mathbf{r}) e^{-2\pi i \mathbf{r} \cdot \mathbf{g}} d\mathbf{r}$ and an inverse Fourier transform is given by $\psi(\mathbf{r}) = \hat{\mathcal{F}}_{\mathbf{g} \rightarrow \mathbf{r}}^{-1} \hat{\psi}(\mathbf{r}) = \int \hat{\psi}(\mathbf{g}) e^{2\pi i \mathbf{r} \cdot \mathbf{g}} d\mathbf{r}$.
 - [22] S. D. Findlay, L. J. Allen, M. P. Oxley, and C. J. Rossouw, Lattice-resolution contrast from a focused coherent electron probe. part II, *Ultramicroscopy* **96**, 65 (2003).
 - [23] D. K. Saldin and P. Rez, The theory of the excitation of atomic inner-shells in crystals by fast electrons, *Philosophical Magazine B* **55**, 481 (1987).
 - [24] M. P. Oxley and L. J. Allen, Delocalization of the effective interaction for inner-shell ionization in crystals, *Physical Review B* **57**, 3273 (1998).
 - [25] Y. Zhu, A. Soukiassian, D. G. Schlom, D. A. Muller, and C. Dwyer, Towards artifact-free atomic-resolution elemental mapping with electron energy-loss spectroscopy, *Applied Physics Letters* **103**, 141908 (2013).
 - [26] A full multislice calculation without these additional approximations is estimated to take 170 years for one frozen phonon pass.